Supplementary Methods

Short Read Preprocessing

Reads are preprocessed differently according to how they will be used: detection of the variant in the tumor, discovery of an artifact in the normal or for variant classification.

For discovery of the variant in the tumor, only the highest quality data should be used in order to eliminate false positives. Therefore only reads that pass the following tests are retained:

(a) Mapping Quality score > 0
(b) Base quality score ≥ 5
(c) If there is an overlapping read pair, and both reads agree the read with the highest quality score is retained otherwise both are discarded.
(d) Sum of the quality scores of the mismatches ≤ 100
(e) < 30% of bases have been soft-clipped
(f) Reads that have not been mapped by "mate rescue" of BWA[1,2] (BAM XT tag ≠ "M")

When looking at the matched normal control to discover systematic artifacts, a less stringent set of filters are applied in order to more readily detect these artifacts. These reads must pass the following tests:

(a) Base quality score ≥ 5
(b) If there is an overlapping read pair, and both reads agree the read with the highest quality score is retained otherwise the read that disagrees with the reference is retained.

Method Parameter Values

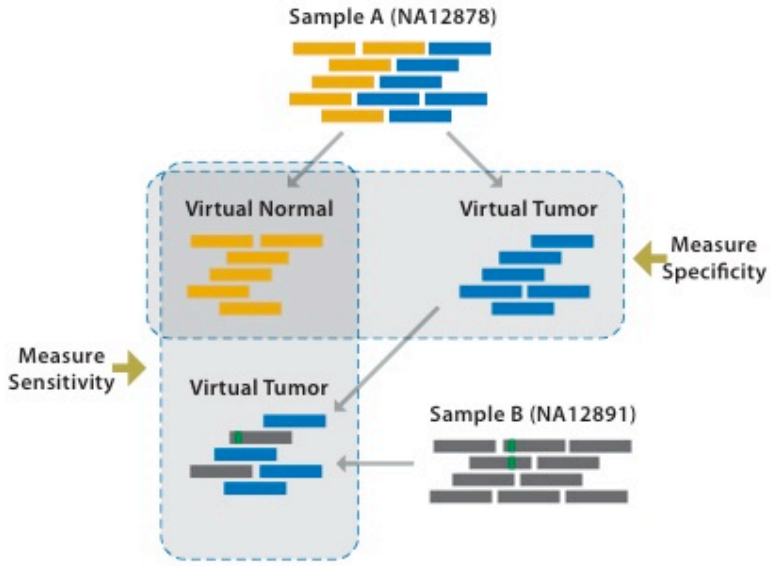
MuTect: v1.1.1 was used for all analysis in this manuscript. The default parameters were used except that *fraction_contamination* was set to 0.0. For the sensitivity measurements using the simulation approach, neither the *dbsnp* nor *panel of normals* were supplied as they would discard all the simulated events.

SomaticSniper[5]: v1.0.0 was run using recommended parameters of *–q 1 –Q 15*. As described in their publication[5], STD results were obtained after filtering initial calls through the *snpfilter.pl.* HC results were obtained by further filtering the output through *fpfilter.pl* and *highconfidence.pl*

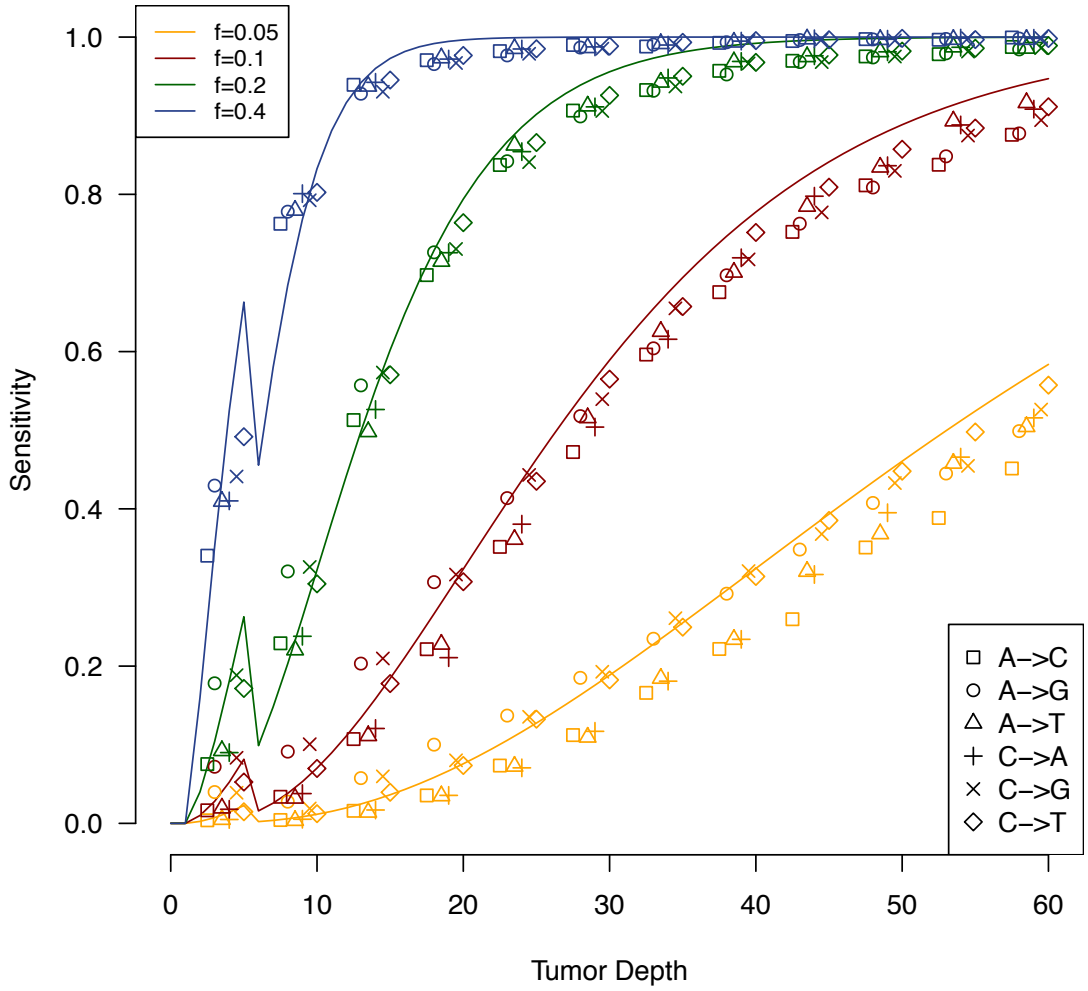
JointSNVMix[6]: v0.7.5 was used, training and then classifying with *joint_snv_mix_two*. In training, *skip_size* was set to 1000 against a tumor and normal BAM. To avoid over-fitting when running on a small number of sites in the downsampling or sensitivity measurements, we trained on a ~10mb region of the same average coverage from the virtual tumor and normal data.

Strelka[7]: v0.4.7 was used with the default parameters for BWA alignments as recommended in the documentation. The depth filter was disabled as the virtual tumor and downsampling BAMs had non-contiguous coverage that would cause this filter to discard most events and unfairly impact sensitivity.

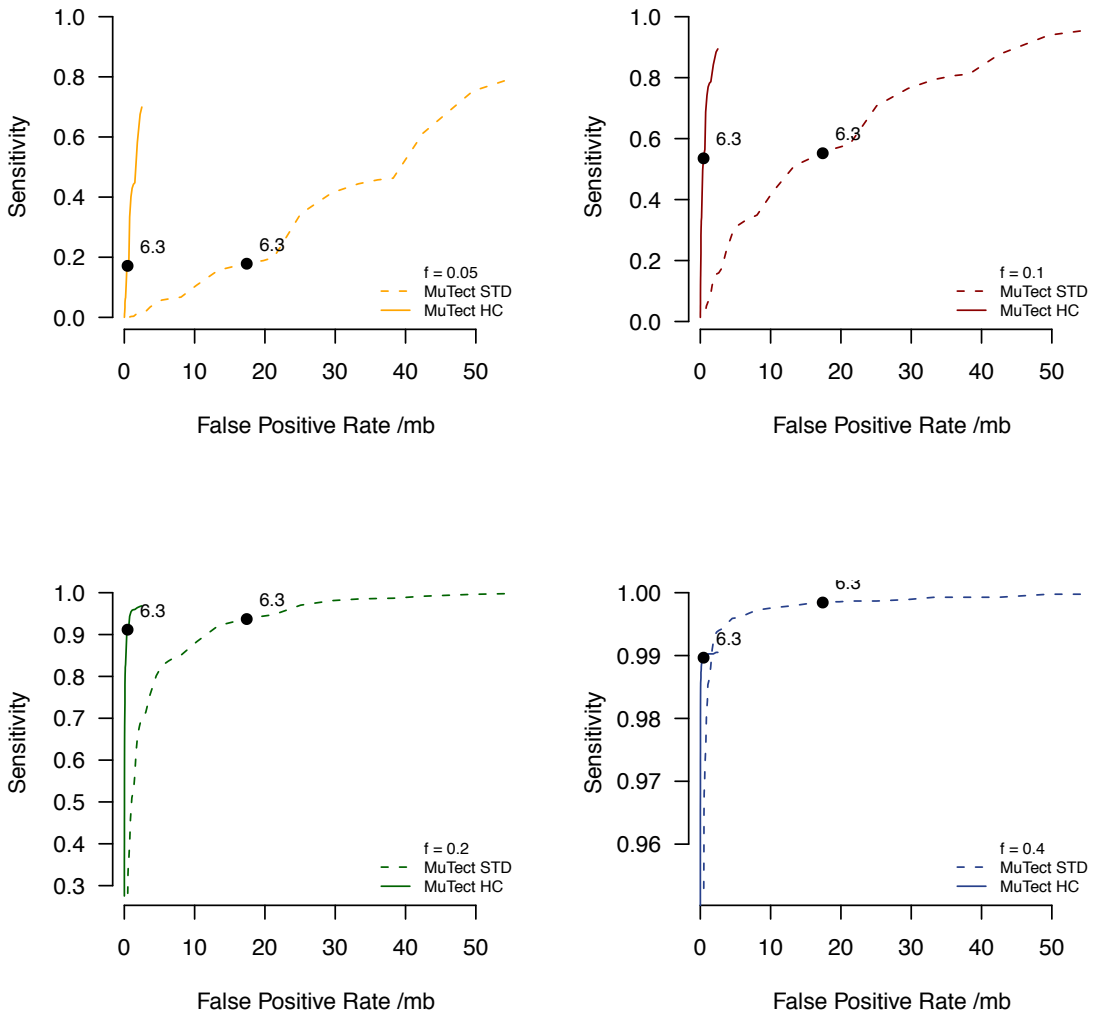
Supplementary Figures



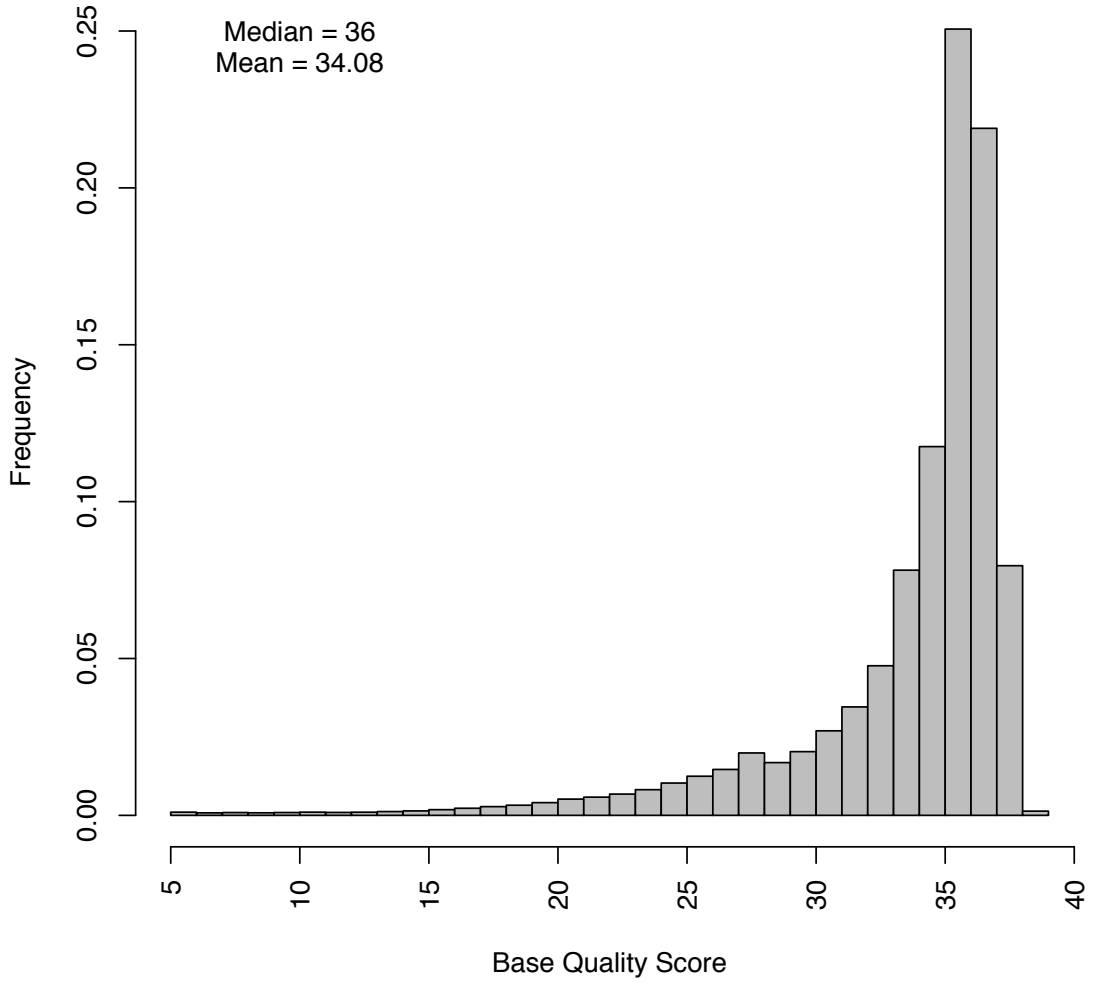
Supplementary Figure 1 | Overview of Virtual Tumor Approach
The Virtual Tumor approach begins with a single sample with multiple libraries (yellow and blue bars). A library is assigned to either a virtual tumor or normal and reads are then randomly drawn from that library to the desired depth. The mutations identified by running a mutation detection method against this tumor-normal pair are false positives and can be used to estimate specificity. Furthermore, a second sample (grey bars) harboring a germline SNP (green) can be used to replace reads in the virtual tumor to simulate a somatic mutation at a controlled allele fraction. Sensitivity of a method can be measured by attempting to detect these events.



Supplementary Figure 2 | Sensitivity by Nucleotide Substitution Type
Sensitivity of MuTect as a function of tumor sequencing depth, mutation allele fraction, and nucleotide substitution as measured by the Virtual Tumor approach. Slightly lower sensitivity is observed for A→C/T and C→A mutations at low alternate allele counts, likely due to lower base quality scores caused by biases in sequencing machine errors.

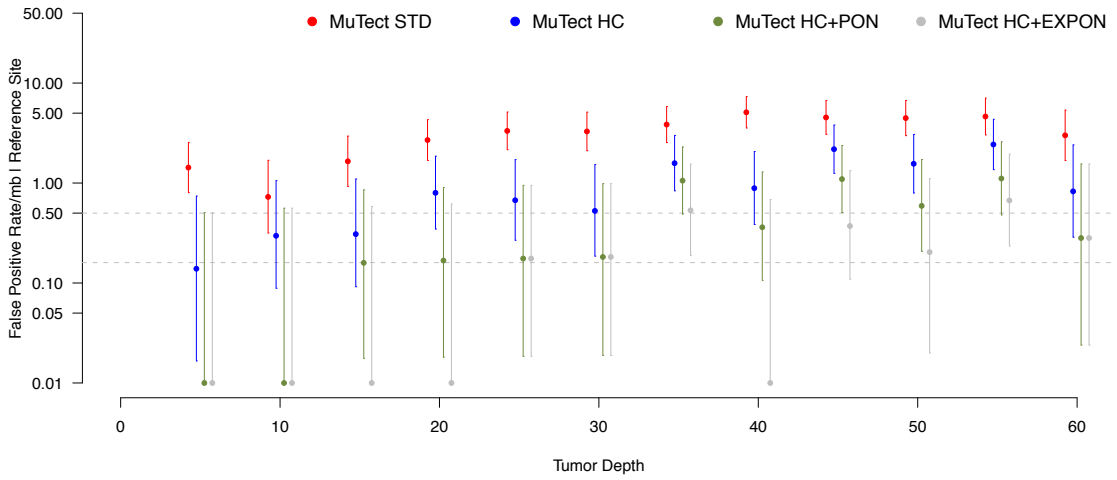


Supplementary Figure 3 | Sensitivity and Specificity of MuTect
The relationship of sensitivity and specificity of MuTect STD (dashed line) and HC (solid line) as measured by the virtual tumors approach with 30x tumor depth and 30x normal depth for several values of allele fraction (yellow = 0.05, red = 0.1, green = 0.2, blue = 0.4) using various values of θ_T. A typical setting of θ_T = 6.3 is marked with black circles.

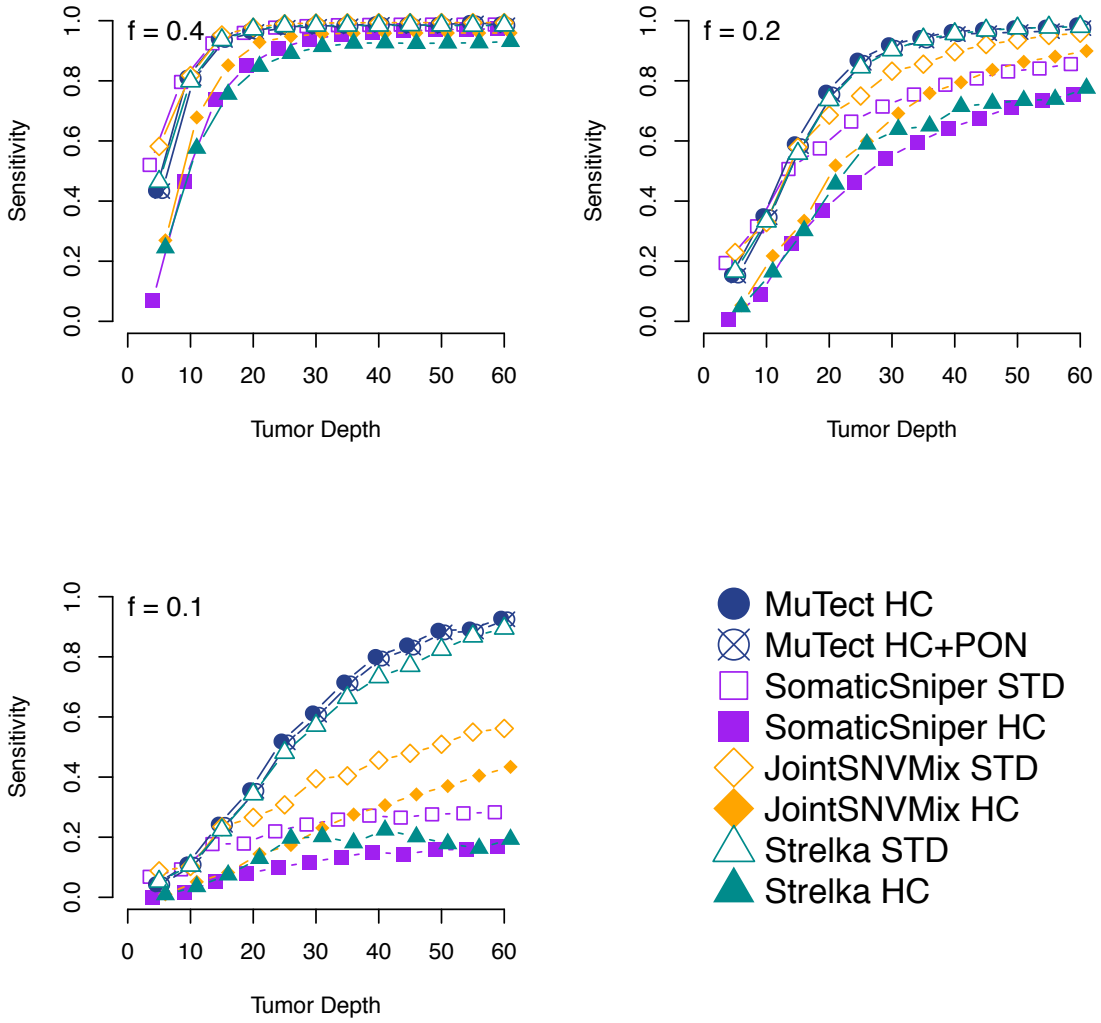


Supplementary Figure 4 | Distribution of Virtual Tumor Simulation Quality Scores

Distribution of base quality score values as observed across a 5mb region of the simulation data set



Supplementary Figure 5 | Specificity in Coding Regions
Somatic miscall error rate for true reference sites in coding regions as a function of tumor sequencing depth for the STD (red) and HC (blue) and HC+PON (green) configurations of MuTect. By using a larger panel of normals (HC+EXPON using 559 1000 Genomes[3] samples) specificity is even higher (grey). Error bars represent 95% CIs. Call sets with no false positives are represented on this log scale as 0.01 errors/Mb. Grey dashed lines represent the calculated error rate (0.5 errors/Mb) and the error rate observed from independent validation (0.16 errors/Mb).



Supplementary Figure 6 | Comparison of Methods using Downsampling Approach

Comparison of sensitivity as a function of tumor sequencing depth and mutation allele fraction for different mutation detection methods and configurations as measured by the downsampling of validated colorectal mutations[8]

References

1. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
2. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics*

25, 2078–2079 (2009).
3. 1000 Genomes Project Consortium A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
4. Gnerre, S. *et al.* High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci USA* **108**, 1513–1518 (2011).
5. Larson, D. E. *et al.* SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics* **28**, 311–317 (2012).
6. Roth, A. *et al.* JointSNVMix: a probabilistic model for accurate detection of somatic mutations in normal/tumour paired next-generation sequencing data. *Bioinformatics* **28**, 907–913 (2012).
7. Saunders, C. T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).
8. Network, T. C. G. A. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).